Noname manuscript No.
(will be inserted by the editor)

Multiscale methods with compactly supported radial basis functions for Galerkin approximation of elliptic PDEs

A. Chernih · Q.T. Le Gia

the date of receipt and acceptance should be inserted later

November 8, 2012

Abstract The aim of this work is to consider multiscale algorithms for solving PDEs with Galerkin methods on bounded domains. We provide results on convergence and condition numbers. We show how to handle PDEs with Dirichlet boundary conditions. We also investigate convergence in terms of the mesh norms and the angles between subspaces to better understand the differences between the algorithms and the observed results. We also consider the issue of the supports of the RBFs overlapping the boundary in our stability analysis, which has not been considered in the literature, to the best of our knowledge.

1 Introduction

Radial basis functions (RBFs) have been increasingly important in the area of approximation theory. For solving partial differential equations (PDEs), RBFs have been studied more for meshless collocation (Giesl & Wendland 2007, Fasshauer 2007), however their use with Galerkin methods have also been considered (Wendland 1998*b*,*c*). Two excellent recent books covering practical and theoretical issues are Fasshauer (2007) and Wendland (2005). A function $\Phi : \mathbb{R}^d \to \mathbb{R}$ is said to be *radial* if there exists a function $\phi : [0, \infty) \to \mathbb{R}$ such

A. Chernih
School of Mathematics and Statistics, University of New South Wales, Sydney NSW 2052, Australia
Tel.: +61-410-697411, Fax: +612 93857123
E-mail: andrew@andrewch.com

Q.T. Le Gia
School of Mathematics and Statistics, University of New South Wales, Sydney NSW 2052, Australia
E-mail: qlegia@unsw.edu.au

that $\Phi(\mathbf{x}) = \phi(\|\mathbf{x}\|_2)$ for all $\mathbf{x} \in \mathbb{R}^d$, where $\|\cdot\|_2$ denotes the usual Euclidean norm in $\mathbb{R}^d$. Then we can define an RBF for a given centre $\mathbf{x}_i \in \mathbb{R}^d$ as

$$\Phi(\mathbf{x}_i) := \phi(\|\mathbf{x} - \mathbf{x}_i\|_2).$$

A practical issue that arises is that of which scale to use for the radial basis functions. A small scale will lead to a sparse and consequently well-conditioned linear system, but at the price of poor approximation power. Conversely, a large scale will have better approximation power but at the price of an ill-conditioned linear system.

Many examples may naturally exhibit multiple scales, for example, constructing an approximation for the height of the earth's surface may suggest a "large scale" to be used over desert regions and a "fine scale" over areas of high variability, such as the Himalayas. It appears much more appropriate to allow different scales in different regions. Of course, this comes at the price of having to select which scales to use in which regions but this is not the topic of this paper.

The multiscale algorithms proposed in this paper are constructed over multiple levels, in which the residual of the current stage is the target function for the next stage, and at each stage, RBFs with smaller support and with more closely spaced centres will be used as basis functions.

Such a multiscale algorithm for interpolation was first proposed in Floater & Iske (1996) and Schaback (1996) but without any theoretical grounding. Theoretical convergence was proven in the case of the data points being located on a sphere (Le Gia et al. 2010) and then extended to interpolation and approximation on bounded domains (Wendland 2010).

In Wendland (1998*c*) we can find various experiments with two multiscale algorithms for constructing Galerkin approximations to PDEs on bounded domains. These two algorithms are studied in this paper. There was no proof of convergence for the first algorithm which we call the *multiscale algorithm*. For the second multiscale algorithm, which we call the *nested multiscale algorithm*, convergence was proven, making use of the fact that the weak formulation of a PDE can be interpreted as a Hilbert space projection method. Numerical experiments with both multiscale algorithms were also given.

The aim of this work is to consider multiscale algorithms for solving PDEs with Galerkin methods on bounded domains. We provide results on convergence and condition numbers. We show how to handle PDEs with Dirichlet boundary conditions which was not covered in the earlier papers. We also investigate convergence in terms of the mesh norms and the angles between subspaces to better understand the differences between the two algorithms and the observed results. We also consider the issue of the supports of the RBFs overlapping the boundary in our stability analysis, which has not been considered in the literature, to the best of our knowledge.

Related multilevel methods for the numerical solution of PDEs include domain decomposition (Smith et al. 1996) and the use of smoothing as a postconditioner with an iterative (multilevel) method (Fasshauer 1999). The heterogeneous multiscale method (Weinan et al. 2007) differs in that it is multiscale

in the time domain, whilst we work with multiple scales in the spatial domain. Multigrid methods (Brenner & Scott 2008) require the use of a computational mesh, whilst our approach is meshfree.

In the next section we review the background material required for the multiscale algorithms. Sections 3 and 4 describe how to construct Galerkin solutions to PDEs using radial basis functions for PDEs without and with Dirichlet boundary conditions respectively. Sections 5 and 6 consider the multiscale and nested multiscale algorithms. Section 7 provides the results of several examples using the two multiscale algorithms. Finally Section 8 provides an analysis of the convergence of the two algorithms.

2 Preliminaries

In this paper, we will use (scaled) compactly supported radial basis functions to construct multiscale approximate solutions to PDEs, that is, we form the solution over multiple levels. We will work with a given domain $\Omega \subseteq \mathbb{R}^d$. A kernel $\Phi : \Omega \times \Omega \to \mathbb{R}$ is also given.

At each level, we will have a finite point set $X \subseteq \Omega$. We will define the *fill distance* as

$$h_{X,\Omega} := \sup_{\mathbf{x} \in \Omega} \min_{\mathbf{x}_j \in X} \|\mathbf{x} - \mathbf{x}_j\|_2,$$

which is a measure of the uniformity of the points in X with respect to Ω. Then for example, at each level i, we denote the fill distance by h_i. The selection of point sets with fill distances decreasing in a specific way will form one of the requirements for convergence of our algorithms.

The functions that we will be concerned with are defined on a bounded domain Ω with a Lipschitz boundary. As a result, there is an extension operator for functions defined in Sobolev spaces which is presented in the following lemma (Brenner & Scott 2008, Theorem 1.4.5). For further details, we refer the reader to Stein (1970).

Lemma 2.1 *Suppose $\Omega \subseteq \mathbb{R}^d$ has a Lipschitz boundary. Then there is an extension mapping $E : H^\tau(\Omega) \to H^\tau(\mathbb{R}^d)$, defined for all non-negative integers τ, satisfying $Ev|_\Omega = v$ for all $v \in H^\tau(\Omega)$ and*

$$\|Ev\|_{H^\tau(\mathbb{R}^d)} \le C\|v\|_{H^\tau(\Omega)}.$$

In this paper, C will denote a generic constant.

Since we also have $\|v\|_{H^\tau(\Omega)} \le \|Ev\|_{H^\tau(\mathbb{R}^d)}$, this means that when we need to consider the $H^\tau(\Omega)$ norms of the errors at each level, we can carry out our error analysis in the $H^\tau(\mathbb{R}^d)$-norm. This is advantageous, since we then have for $g \in H^\tau(\mathbb{R}^d)$

$$\|g\|^2_{H^\tau(\mathbb{R}^d)} = \int_{\mathbb{R}^d} |\widehat{g}(\boldsymbol{\omega})|^2 \left(1 + \|\boldsymbol{\omega}\|_2^2\right)^\tau \mathrm{d}\boldsymbol{\omega}, \tag{1}$$

upon defining the Fourier transform as

$$\widehat{g}(\boldsymbol{\omega}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} g(\boldsymbol{x}) e^{-i\mathbf{x}^T\boldsymbol{\omega}} \mathrm{d}\mathbf{x}.$$

At each level, we will also require a scaled version of the kernel $\Phi : \Omega \times \Omega \to \mathbb{R}$. For our unscaled kernel we will use a Wendland compactly supported radial basis function (Wendland 2005). With a (level-specific) scaling parameter $\delta > 0$, we can define the scaled kernels as

$$\Phi_\delta(\mathbf{x}, \mathbf{y}) := \delta^{-d} \phi \left(\frac{\|\mathbf{x} - \mathbf{y}\|_2}{\delta} \right). \tag{2}$$

Appropriate selection of the scaling parameters will also prove to be one of the important ingredients for convergence of our multiscale algorithms.

For the Wendland basis functions, there exist two constants $0 < c_1 \leq c_2$ such that their Fourier transforms satisfy (Wendland 2005)

$$c_1 \left(1 + \|\boldsymbol{\omega}\|_2^2\right)^{-\tau} \leq \widehat{\Phi}(\boldsymbol{\omega}) \leq c_2 \left(1 + \|\boldsymbol{\omega}\|_2^2\right)^{-\tau}, \quad \boldsymbol{\omega} \in \mathbb{R}^d, \tag{3}$$

where $\tau = (d + 2k + 1)/2$ where d is the spatial dimension and k is the smoothness parameter. For further details, we refer the reader to Wendland (2005). The native space $\mathcal{N}_\Phi(\mathbb{R}^d)$ of Φ consists of all functions $g \in L_2(\mathbb{R}^d)$ such that

$$\|g\|_\Phi^2 = \int_{\mathbb{R}^d} \frac{|\widehat{g}(\boldsymbol{\omega})|^2}{\widehat{\Phi}(\boldsymbol{\omega})} \mathrm{d}\boldsymbol{\omega} < \infty. \tag{4}$$

Taking (1) and (3) together shows that $\mathcal{N}_\Phi(\mathbb{R}^d)$ is norm-equivalent to the Sobolev space $H^\tau(\mathbb{R}^d)$.

Consequently the Fourier transform of Φ_δ, $\widehat{\Phi_\delta}(\boldsymbol{\omega}) = \widehat{\Phi}(\delta\boldsymbol{\omega})$, satisfies

$$c_1 \left(1 + \delta^2\|\boldsymbol{\omega}\|_2^2\right)^{-\tau} \leq \widehat{\Phi_\delta}(\boldsymbol{\omega}) \leq c_2 \left(1 + \delta^2\|\boldsymbol{\omega}\|_2^2\right)^{-\tau}, \quad \boldsymbol{\omega} \in \mathbb{R}^d. \tag{5}$$

We will need norm equivalence as stated in the following lemma.

Lemma 2.2 *For every $\delta \in (0, \delta_a]$ and for all $g \in H^\tau(\mathbb{R}^d)$, there exist constants $0 < c_3 \leq c_4$ such that*

$$c_3 \|g\|_{\Phi_\delta} \leq \|g\|_{H^\tau(\mathbb{R}^d)} \leq c_4 \delta^{-\tau} \|g\|_{\Phi_\delta}.$$

Proof The case $\delta_a \leq 1$ was proven in Wendland (2010) with $c_3 = c_1^{1/2}$ and $c_4 = c_2^{1/2}$. To extend this to the case where $\delta_a > 1$, note that for $\delta > 1$ we have

$$\left(1 + \|\boldsymbol{\omega}\|_2^2\right)^\tau = \delta^{-2\tau} \left(\delta^2 + \delta^2\|\boldsymbol{\omega}\|_2^2\right)^\tau \geq \delta_a^{-2\tau} \left(1 + \delta^2\|\boldsymbol{\omega}\|_2^2\right)^\tau.$$

Together with (1), (5) and (4) we can see that

$$\begin{aligned}
\|g\|^2_{H^\tau(\mathbb{R}^d)} &= \int\limits_{\mathbb{R}^d} |\widehat{g}(\boldsymbol{\omega})|^2 \left(1+\|\boldsymbol{\omega}\|_2^2\right)^\tau \mathrm{d}\boldsymbol{\omega} \\
&\geq \delta_a^{-2\tau} \int\limits_{\mathbb{R}^d} |\widehat{g}(\boldsymbol{\omega})|^2 \left(1+\delta^2\|\boldsymbol{\omega}\|_2^2\right)^\tau \mathrm{d}\boldsymbol{\omega} \\
&\geq c_1\delta_a^{-2\tau} \int\limits_{\mathbb{R}^d} \frac{|\widehat{g}(\boldsymbol{\omega})|^2}{\widehat{\Phi_\delta}(\boldsymbol{\omega})} \mathrm{d}\boldsymbol{\omega} \\
&\geq c_1\delta_a^{-2\tau}\|g\|^2_{\Phi_\delta}.
\end{aligned}$$

For the lower bound, we can just use $\delta > 1$ directly to derive

$$\begin{aligned}
\|g\|^2_{H^\tau(\mathbb{R}^d)} &= \int\limits_{\mathbb{R}^d} |\widehat{g}(\boldsymbol{\omega})|^2 \left(1+\|\boldsymbol{\omega}\|_2^2\right)^\tau \mathrm{d}\boldsymbol{\omega} \\
&\leq \int\limits_{\mathbb{R}^d} |\widehat{g}(\boldsymbol{\omega})|^2 \left(1+\delta^2\|\boldsymbol{\omega}\|_2^2\right)^\tau \mathrm{d}\boldsymbol{\omega} \\
&\leq c_2\|g\|^2_{\Phi_\delta},
\end{aligned}$$

using (5) and (4). Then setting $c_3 := c_1^{1/2}\min(1,\delta_a^{-\tau})$ and $c_4 := c_2^{1/2}\max(1,\delta_a^{\tau})$ completes the proof. □

3 PDEs with Neumann and/or Robin boundary conditions

In this section, we consider a second order PDE which has homogeneous Neumann and/or Robin boundary conditions. For example, such a PDE with Neumann boundary conditions is given by

$$\mathcal{L}u = f \quad \text{in } \Omega, \tag{6a}$$

$$\frac{\partial u}{\partial \mathbf{n}} = 0 \quad \text{on } \partial\Omega, \tag{6b}$$

where $\mathcal{L}$ is a second order elliptic differential operator, $\mathbf{n}$ denotes the outward unit normal vector and $\partial\Omega$ denotes the boundary. The weak formulation is given by

$$a(u,v) = \langle f, v\rangle_{L_2(\Omega)} \quad \forall\, v \in V, \tag{7}$$

where $V = H^1(\Omega)$. We assume that $\mathcal{L}$ and f are such that $a(u,v)$ is a strictly coercive and continuous bilinear form defined on $V \times V$ and $\langle f, v\rangle_{L_2(\Omega)}$ is a continuous linear form defined on V. By the Lax-Milgram theorem, (7) has a unique solution $u \in V$. We will also require $u \in H^2(\Omega)$ with spatial dimension $d \leq 3$.

Galerkin approximation seeks to solve (7) with a finite dimensional subspace $V_N \subseteq V$. In other words, the Galerkin approximation $\widetilde{u}_N$ is the solution of

$$\widetilde{u}_N \in V_N : a(\widetilde{u}_N, v) = \langle f, v \rangle_{L_2(\Omega)} \quad \forall\, v \in V_N. \tag{8}$$

We will consider Ω to be a bounded domain with a Lipschitz boundary, which means that we can apply the extension operator to use norms in $\mathbb{R}^d$ as explained in the introduction. For further information on weak formulation of PDEs and Galerkin approximation, we refer the reader to Brenner & Scott (2008).

Since the PDE does not have Dirichlet boundary conditions, we can use the entire Sobolev space $H^1(\Omega)$ rather than the subspace $\mathring{H}^1(\Omega)$ consisting of functions with zero boundary values which can occur with pure Dirichlet boundary conditions.

We will consider finite dimensional subspaces $V_N \subseteq V$ of the form

$$V_N := \operatorname{span}\{\Phi(\cdot, \mathbf{x}_j) : 1 \le j \le N\}$$

where $\Phi : \mathbb{R}^d \to \mathbb{R}$ is at least a C^1-function and there are N centres $\{\mathbf{x}_j : 1 \le j \le N\}$. In this case our approximation takes the form

$$\widetilde{u}_N = \sum_{j=1}^{N} c_j \Phi(\cdot, \mathbf{x}_j),$$

and the weak formulation with this approximation given by

$$a(\widetilde{u}_N, v) = \langle f, v \rangle_{L_2(\Omega)} \quad \forall\, v \in V_N,$$

results in a linear system

$$\mathbf{A}\mathbf{c} = \mathbf{f}$$

where the entries of the *stiffness matrix* are given by

$$A_{ij} = a(\Phi(\cdot, \mathbf{x}_i), \Phi(\cdot, \mathbf{x}_j)) \tag{9}$$

and

$$\mathbf{f}_i = \int_\Omega f(\mathbf{x}) \Phi(\mathbf{x}, \mathbf{x}_i)\, \mathrm{d}\mathbf{x}.$$

We have the following result from Wendland (1998*b*).

Theorem 3.1 *If $u \in H^2(\Omega), d \le 3$, is the solution to the variational problem* (7) *and $\widetilde{u}_N \in V_N$ is the solution of* (8)*, where V_N is generated with X satisfying $h \le h_0$ for h_0 small enough and fixed, then the error can be bounded by*

$$\|u - \widetilde{u}_N\|_{H^1(\Omega)} \le C h \|u\|_{H^2(\Omega)}.$$

Lemma 3.2 *Consider the adjoint variational problem*

$$a(v, u) = \langle f, v \rangle_{L_2(\Omega)} \quad \forall\, v \in V. \tag{10}$$

If the solution u satisfies the regularity estimate

$$\|u\|_{H^2(\Omega)} \leq C\|f\|_{L_2(\Omega)} \tag{11}$$

for both the variational problem (7) *and the adjoint variational problem* (10)*, then we have the following error bound*

$$\|u - \widetilde{u}_N\|_{L_2(\Omega)} \leq Ch\|u - \widetilde{u}_N\|_{H^1(\Omega)} \leq Ch^2\|u\|_{H^2(\Omega)}.$$

Proof We follow a duality argument as in (Brenner & Scott 2008, Theorem 5.7.6). With $\widetilde{u}_N$ as defined in (8), let w be the solution to the adjoint problem

$$a(v, w) = \langle u - \widetilde{u}_N, v \rangle_{L_2(\Omega)} \quad \forall\, v \in V,$$

and let the Galerkin approximation be given by $\widetilde{w}$. Then since the bilinear form $a(\cdot, \cdot)$ is bounded, $u \in H^2(\Omega)$, and with Theorem 3.1 we have

$$\begin{aligned}
\|u - \widetilde{u}_N\|^2_{L_2(\Omega)} &= \langle u - \widetilde{u}_N, u - \widetilde{u}_N \rangle_{L_2(\Omega)} \\
&= a(u - \widetilde{u}_N, w) = a(u - \widetilde{u}_N, w - \widetilde{w}) \\
&\leq C\|u - \widetilde{u}_N\|_{H^1(\Omega)}\, \|w - \widetilde{w}\|_{H^1(\Omega)} \\
&\leq Ch\|u - \widetilde{u}_N\|_{H^1(\Omega)}\, \|w\|_{H^2(\Omega)} \\
&\leq Ch\|u - \widetilde{u}_N\|_{H^1(\Omega)}\, \|u - \widetilde{u}_N\|_{L_2(\Omega)},
\end{aligned}$$

where in the second last line, we use the regularity estimate (11) and the result follows with another application of Theorem 3.1. □

We note that (11) is known to hold (Brenner & Scott 2008, p.139)

- if Ω has a smooth boundary and the problem has pure Dirichlet or pure Neumann boundary conditions;
- if $d = 2$ and Ω is convex and the problem has pure Dirichlet or pure Neumann boundary conditions.

4 PDEs with Dirichlet boundary conditions

In this section we will consider a PDE with Dirichlet boundary conditions and we seek error estimates similar to those in the previous section. For ease of description, we will consider the following boundary value problem

$$-\Delta u = f \quad \text{in } \Omega, \tag{12a}$$

$$u = g \quad \text{on } \partial\Omega, \tag{12b}$$

where Δ denotes the Laplacian in $\mathbb{R}^d$.

Nitsche (1970–1971) proposed minimising the functional $J(v - u)$ where

$$J(w) := \int_{\Omega} |\nabla w|^2 - 2 \int_{\partial\Omega} w\,(\nabla w \cdot \mathbf{n}) + \beta_N \int_{\partial\Omega} w^2,$$

for all $v \in V_N$ with u being the solution of (12) and where $\mathbf{n}$ denotes the outward unit normal vector and ∇ the gradient operator. The parameter $\beta_N > 0$ depends only on the subspace V_N. The approximation $\widetilde{u}_N$ is given by $J(\widetilde{u}_N - u) := \inf_{v \in V_N} J(v - u)$. With f and g from (12), we can compute $\widetilde{u}_N$ since

$$J(v-u) = J(v) + J(u) - 2\left(\int_{\Omega} fv + \int_{\partial\Omega} g\,(\beta_N v - \nabla v \cdot \mathbf{n}) \right).$$

Then the variational form to approximate (12) becomes: find $\widetilde{u}_N \in V_N$ such that for all $v \in V_N$

$$a_D(\widetilde{u}_N, v) = \ell_D(v), \tag{13}$$

where we use the subscript D to denote the Dirichlet boundary conditions and

$$a_D(u,v) := \int_{\Omega} \nabla u \cdot \nabla v - \int_{\partial\Omega} v\,(\nabla u \cdot \mathbf{n}) - \int_{\partial\Omega} u\,(\nabla v \cdot \mathbf{n}) + \beta_N \int_{\partial\Omega} uv \tag{14a}$$

$$\ell_D(v) := \int_{\Omega} fv - \int_{\partial\Omega} g\,(\nabla v \cdot \mathbf{n}) + \beta_N \int_{\partial\Omega} vg. \tag{14b}$$

It can be shown that the variational form using Nitsche's method leads to variational consistency, in the sense that if u is sufficiently regular, then (Blowey et al. 2003, p. 119)

$$a_D(u,v) = \ell_D(v), \quad \forall\, v \in V_N.$$

If there exists a positive constant C_N such that

$$\|\nabla v \cdot \mathbf{n}\|_{L_2(\partial\Omega)} \le \frac{C_N}{\sqrt{\delta}} \|\nabla v\|_{L_2(\Omega)} \quad \forall\, v \in V_N, \tag{15}$$

with δ being the support of the radial basis functions, then selecting

$$\beta_N = \frac{c_5}{\delta} \tag{16}$$

with $c_5 > 2C_N^2$ will ensure that the bilinear form $a_D(\cdot,\cdot)$ is symmetric positive definite. This choice of c_5 will also ensure that a_D is coercive since

$$\begin{aligned}
a_D(v,v) &= \|\nabla v\|^2_{L_2(\Omega)} - 2 \int_{\partial\Omega} v\,(\nabla v \cdot \mathbf{n}) + \beta_N \|v\|^2_{L_2(\partial\Omega)} \\
&\ge \|\nabla v\|^2_{L_2(\Omega)} - \frac{2C_N}{\delta^{1/2}} \|v\|_{L_2(\partial\Omega)} \|\nabla v\|_{L_2(\Omega)} + \beta_N \|v\|^2_{L_2(\partial\Omega)} \\
&\ge \frac{1}{2} \|\nabla v\|^2_{L_2(\Omega)} + \left(\beta_N - \frac{2C_N^2}{\delta} \right) \|v\|^2_{L_2(\partial\Omega)} \\
&\ge C \|v\|^2_{H^1(\Omega)}, \forall\, v \in V_N
\end{aligned}$$

where we have used the Cauchy-Schwarz and Friedrichs inequalities, (15) and that $2xy \leq x^2 + y^2$. We recall that the Friedrichs inequality (Maz'ya 2011) states that

$$\int\limits_{\Omega} |u|^2 \leq C \left(\int\limits_{\Omega} |\nabla u|^2 + \int\limits_{\partial\Omega} |u|^2 \right),$$

for Ω being a bounded domain for which the Gauss-Green formula holds.

Continuity follows since

$$\begin{aligned} |a_D(u,v)| &\leq |u|_{H^1(\Omega)}|v|_{H^1(\Omega)} + C_N/\delta \left(\|v\|_{L_2(\partial\Omega)}\|\nabla u\|_{L_2(\Omega)} + \right. \\ &\quad \left. \|u\|_{L_2(\partial\Omega)}\|\nabla v\|_{L_2(\Omega)} \right) + \beta_N \|u\|_{L_2(\partial\Omega)}\|v\|_{L_2(\partial\Omega)} \\ &\leq |u|_{H^1(\Omega)}|v|_{H^1(\Omega)} + C \left(\|v\|_{L_2(\Omega)}\|\nabla u\|_{L_2(\Omega)} + \right. \\ &\quad \left. \|u\|_{L_2(\Omega)}\|\nabla v\|_{L_2(\Omega)} \right) + \beta_N \|u\|_{L_2(\Omega)}\|v\|_{L_2(\Omega)} \\ &\leq C\|u\|_{H^1(\Omega)}\|v\|_{H^1(\Omega)}, \quad \forall\, u, v \in V_N \end{aligned}$$

where we have used the Cauchy-Schwarz inequality, (15) and the Sobolev trace embedding theorem.

Nitsche also proved that the optimal error estimates of Theorem 3.1 and Lemma 3.2 hold in this setting if, in addition to the requirement of selecting β_N satisfying (16), there exists a $s_u \in V_N$ such that for $u \in H^2(\Omega)$, the following error bounds hold for $k \in \{0,1\}$

$$\|u - s_u\|_{H^k(\Omega)} \leq Ch^{2-k}\|u\|_{H^2(\Omega)}, \tag{17a}$$

$$\|u - s_u\|_{H^k(\partial\Omega)} \leq Ch^{3/2-k}\|u\|_{H^2(\Omega)}. \tag{17b}$$

With our choice of compactly supported radial basis functions, this requirement is known to hold (Wendland 1998*b*).

Note that the most challenging aspect of Nitsche's method is the derivation of the weak form and the selection of the stabilisation parameter β_N. Both the weak form and the choice of the parameter β_N depend on the PDE as well as the Dirichlet boundary conditions.

5 Multiscale Galerkin approximation

In this section, we consider a multiscale algorithm for constructing a Galerkin approximation where we use the residual from the previous level as the target for each subsequent level. We define the approximation at level i as $\widetilde{u}_i := \widetilde{u}_{N_i}$ with centres N_i and the approximation space at level i as $V_i := V_{N_i}$. The algorithm is given in Algorithm 1. The bilinear form $a(\cdot,\cdot)$ used in this algorithm is the unmodified bilinear form in the case of a PDE with Neumann or Robin boundary conditions and the Nitsche's method bilinear form $a_D(\cdot,\cdot)$ in the case of a PDE with Dirichlet boundary conditions.

The algorithm as stated uses the same bilinear form at each level and it is the approximation space V_i which changes. However the Nitsche's method

Algorithm 1: Multiscale Galerkin approximation

Data: n: number of levels
$\{X_i\}_{i=1}^n$: the set of nested centres for each level i, with mesh norms at each level given by h_i satisfying $c\mu h_i \le h_{i+1} \le \mu h_i$ with fixed $\mu \in (0,1), c \in (0,1]$ and h_1 sufficiently small
$\{\delta_i\}_{i=1}^n$: the scale parameters to use at each level, satisfying $\delta_i = \nu h_i$, ν a fixed constant.

begin
Set $\widetilde{u}_0 = 0$
for $i = 1, 2, \ldots, n$ **do**
With the level-specific approximation subspace $V_i := \text{span}\{\Phi_{\delta_i}(\cdot, \mathbf{x}), \mathbf{x} \in X_i\}$ solve the Galerkin approximation given by

$$\text{Find } s_i \in V_i \,:\, a(s_i, v) = \langle f, v \rangle_{L_2(\Omega)} - a(\widetilde{u}_{i-1}, v) \;\forall v \in V_i$$

Update the solution according to

$$\widetilde{u}_i = \widetilde{u}_{i-1} + s_i$$

Result: Approximate solution at level n, $\widetilde{u}_n$
The error at level n, $e_n := u - \widetilde{u}_n$.

bilinear form $a_D(\cdot,\cdot)$ will vary at each level since the value of C_N^2 is proportional to δ^{-1} and hence so is β_N. This means that we will need to select the value of β_N corresponding to the last level and to use this for all previous levels. This will also mean that we will need to know the number of levels in advance.

Henceforth we will simply refer to the bilinear form as $a(\cdot,\cdot)$. This should cause no confusion as we have the same error bounds in both cases, as well as coercivity and continuity, and the multiscale algorithm follows the same steps in both cases. We require one more lemma before we can analyse the convergence of the multiscale algorithm.

Lemma 5.1 *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain with a Lipschitz boundary. Then for Algorithm 1 and for a given level $i > 1$, we have the following bound on the $H^1(\Omega)$ error between subsequent levels.*

$$\|e_i\|_{H^1(\Omega)} \le C\|e_{i-1}\|_{H^1(\Omega)},$$

where e_i is defined in Algorithm 1.

Proof We will firstly show that s_i is the Galerkin approximation of e_{i-1}. We have

$$\begin{aligned} a(s_i, w) &= \langle f, w\rangle_{L_2(\Omega)} - a(\widetilde{u}_{i-1}, w), \; w \in V_i \\ &= a(u, w) - a(\widetilde{u}_{i-1}, w) \\ &= a(u - \widetilde{u}_{i-1}, w) \\ &= a(e_{i-1}, w), \end{aligned}$$

where we have used the variational form of the PDE and the linearity in the first argument of the bilinear form $a(\cdot, \cdot)$. Hence on setting $w = s_i$ we obtain

$$a(e_{i-1} - s_i, s_i) = 0.$$

Upon noting that $e_i = e_{i-1} - s_i$, it follows easily that

$$a(e_i, e_i) = a(e_{i-1}, e_{i-1}) - a(s_i, s_i),$$

and since the bilinear form a is continuous and coercive

$$\|e_i\|^2_{H^1(\Omega)} + \|s_i\|^2_{H^1(\Omega)} \le C\|e_{i-1}\|^2_{H^1(\Omega)},$$

from which the result follows. □

The following theorem and corollaries are our main results for the convergence of the multiscale Galerkin approximation. For the error analysis, we will need the norm

$$\|u\|^2_{\Psi_j} := \int_{\mathbb{R}^d} |\widehat{u}(\boldsymbol{\omega})|^2 \left(1 + \delta_j^2 \|\boldsymbol{\omega}\|_2^2\right) \mathrm{d}\boldsymbol{\omega}.$$

As in Lemma 2.2, this norm satisfies

$$c_7\|u\|_{\Psi_j} \le \|u\|_{H^1(\mathbb{R}^d)} \le c_8 \delta_j^{-1} \|u\|_{\Psi_j}. \tag{18}$$

Theorem 5.2 *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain with a Lipschitz boundary. Then for Algorithm 1 there exists a constant $\alpha_1 > 0$ such that*

$$\|Ee_i\|_{\Psi_{i+1}} \le \alpha_1 \|Ee_{i-1}\|_{\Psi_i} \quad \text{for} \;\; i = 1, 2, \ldots$$

where Ee_i is the extension operator defined in Lemma 2.1 applied to e_i. The constant α_1 satisfies $\alpha_1 < 1$ if in Algorithm 1 ν is sufficiently small and μ is sufficiently large.

Proof Using (4) and (5), we can write

$$\|Ee_i\|^2_{\Psi_{j+1}} \le \frac{1}{c_1} \int_{\mathbb{R}^d} |\widehat{Ee_i}(\boldsymbol{\omega})|^2 \left(1 + \delta_{i+1}^2 \|\boldsymbol{\omega}\|_2^2\right) \mathrm{d}\boldsymbol{\omega} =: \frac{1}{c_1}\left(I_1 + I_2\right)$$

with

$$I_1 := \int_{\|\boldsymbol{\omega}\|_2 \le \frac{1}{\delta_{i+1}}} |\widehat{Ee_i}(\boldsymbol{\omega})|^2 \left(1 + \delta_{i+1}^2 \|\boldsymbol{\omega}\|_2^2\right) \mathrm{d}\boldsymbol{\omega},$$

$$I_2 := \int\limits_{\|\boldsymbol{\omega}\|_2 \geq \frac{1}{\delta_{i+1}}} |\widehat{Ee_i}(\boldsymbol{\omega})|^2 \left(1 + \delta_{i+1}^2 \|\boldsymbol{\omega}\|_2^2\right) \mathrm{d}\boldsymbol{\omega}.$$

Now we consider the first integral where we can use that $\delta_{i+1}\|\boldsymbol{\omega}\|_2 \leq 1$ and then Lemmas 3.2 and (18).

$$\begin{aligned}
I_1 &\leq 2 \int\limits_{\|\boldsymbol{\omega}\|_2 \leq \frac{1}{\delta_{i+1}}} |\widehat{Ee_i}(\boldsymbol{\omega})|^2 \, \mathrm{d}\boldsymbol{\omega} \\
&\leq 2 \, \|Ee_i\|_{L_2(\mathbb{R}^d)}^2 \leq C \, \|e_i\|_{L_2(\Omega)}^2 \leq C h_i^2 \, \|e_i\|_{H^1(\Omega)}^2 \\
&\leq C \, h_i^2 \, \|e_{i-1}\|_{H^1(\Omega)}^2 \leq C \left(\frac{h_i}{\delta_i}\right)^2 \|Ee_{i-1}\|_{\Psi_i}^2 \\
&= C\nu^{-2} \, \|Ee_{i-1}\|_{\Psi_i}^2,
\end{aligned}$$

where we have also used Lemma 5.1. For I_2, since $\delta_{i+1}\|\boldsymbol{\omega}\|_2 \geq 1$, we have that

$$\left(1 + \delta_{i+1}^2 \|\boldsymbol{\omega}\|_2^2\right) \leq 2\delta_{i+1}^2 \|\boldsymbol{\omega}\|_2^2 \leq 2\delta_{i+1}^2 \left(1 + \|\boldsymbol{\omega}_2^2\right).$$

Then again using Lemma 5.1 shows that

$$\begin{aligned}
I_2 &\leq 2\delta_{i+1}^2 \|Ee_i\|_{H^1(\mathbb{R}^d)}^2 \leq C\delta_{i+1}^2 \|e_i\|_{H^1(\Omega)}^2 \\
&\leq C\delta_{i+1}^2 \|e_{i-1}\|_{H^1(\Omega)}^2 \leq C \left(\frac{\delta_{i+1}}{\delta_i}\right)^2 \|Ee_{i-1}\|_{\Psi_i}^2 \\
&\leq C\mu^2 \|Ee_{i-1}\|_{\Psi_i}^2.
\end{aligned} \tag{19}$$

Combining our results for I_1 and I_2 and now writing C_1 and C_2 for the two constants appearing in the bounds of the expressions for I_1 and I_2 respectively, we have that

$$\|Ee_i\|_{\Psi_{i+1}}^2 \leq \left(\nu^{-2} C_1/c_1 + \mu^2 C_2/c_1\right) \|Ee_{i-1}\|_{\Psi_i}^2,$$

and the result follows with

$$\alpha_1 := \left(\nu^{-2} C_1/c_1 + \mu^2 C_2/c_1\right)^{1/2}. \tag{20}$$

□

Corollary 5.3 *There exists a constant $C_3 > 0$ such that*

$$\|u - \widetilde{u}_n\|_{L_2(\Omega)} \leq C_3 \alpha_1^n \|u\|_{H^1(\Omega)} \quad \textit{for} \;\; n = 1, 2, \ldots \tag{21}$$

Thus $\widetilde{u}_n$ resulting from Algorithm 1 converges linearly to u in the L_2-norm if $\alpha_1 < 1$.

Proof Using Lemmas 3.2 and 2.2 again, we can see that

$$\begin{aligned}\|u-\widetilde{u}_n\|_{L_2(\Omega)} &= \|e_n\|_{L_2(\Omega)} \le Ch_n\|e_n\|_{H^1(\Omega)} \\ &\le Ch_n\|Ee_n\|_{H^1(\mathbb{R}^d)} \le C\,h_n\,\delta_{n+1}^{-1}\|Ee_n\|_{\Psi_{n+1}} \\ &\le C\|Ee_n\|_{\Psi_{n+1}},\end{aligned}$$

since

$$\frac{h_n}{\delta_{n+1}} = \frac{h_n}{\nu h_{n+1}} \le \frac{1}{c\mu\nu}.$$

Now we can apply Theorem 5.2 n times, and noting that $\widetilde{u}_0 = 0$, leads to

$$\|u-\widetilde{u}_n\|_{L_2(\Omega)} \le C\alpha_1^n\|Eu\|_{\Psi_1} \le C\alpha_1^n\|Eu\|_{H^1(\mathbb{R}^d)} \le C\alpha_1^n\|u\|_{H^1(\Omega)}.$$

□

5.1 Condition number

In this subsection, we present upper and lower bounds for the eigenvalues of the multiscale Galerkin algorithm. Since the Galerkin approximation matrix is symmetric and positive definite, we know that the condition number is given by

$$\kappa(\mathbf{A}) = \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})}, \tag{22}$$

where $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ denote the maximum and minimum eigenvalues of $\mathbf{A}$ as defined in (9).

Theorem 5.4 *Let Φ be a positive definite kernel generating $H^\tau(\mathbb{R}^d)$ with $\tau > d/2$. Let $\Phi_i := \Phi(\cdot, \mathbf{x}_i)$ and assume that there exists a constant $c_6 > 0$ depending only on Φ and Ω such that*

$$\alpha^T\left(\mathbf{F} - c_6\mathbf{G}\right)\alpha \ge 0, \quad \forall\ \alpha \in \mathbb{R}^N, \tag{23}$$

which means that $\mathbf{F} - c_6\mathbf{G}$ is positive semi-definite and where

$$\begin{aligned}F_{i,j} &= \left\langle \Phi_i, \Phi_j \right\rangle_{H^1(\Omega)}, \\ G_{i,j} &= \left\langle \Phi_i, \Phi_j \right\rangle_{H^1(\mathbb{R}^d)}.\end{aligned}$$

Then the condition number of $\mathbf{A}$ can be bounded by

$$\kappa(\mathbf{A}) \le C\left(\frac{1}{q_X}\right)^{4\tau-2d},$$

where the constant C is independent of the point set X.

Proof Since $v = \sum_{i=1}^{N} \alpha_i \Phi_i$ and with with the coercivity of $a(\cdot,\cdot)$ and (23), we have that

$$\begin{aligned} a(v,v) &\geq C\|v\|^2_{H^1(\Omega)} = C\alpha^T \mathbf{F} \alpha^T \\ &\geq C\alpha^T \mathbf{G} \alpha \geq C\lambda_{\min}(\mathbf{G}) \|\alpha\|_2^2, \end{aligned}$$

where $\lambda_{\min}(\mathbf{G})$ is the minimum eigenvalue of $\mathbf{G}$. From Wendland (1998*c*), we know that $\langle \Phi_i, \Phi_j \rangle_{H^1(\mathbb{R}^d)}$ is a radial function given by

$$\Upsilon(\mathbf{x}_i, \mathbf{x}_j) := -(\Delta\Phi) \star \Phi(\mathbf{x}_i - \mathbf{x}_j) + \Phi \star \Phi(\mathbf{x}_i - \mathbf{x}_j),$$

where Δ again denotes the Laplace operator and $\star$ denotes convolution defined as $f \star g(x) := \int f(y) g(x-y)\, \mathrm{d}y$. From (Wendland 2005, Theorem 12.3) we know that we can use $\widehat{\Upsilon}$ to derive a lower bound on the minimum eigenvalue of $\mathbf{G}$. Then we have that

$$\widehat{\Upsilon}(\mathbf{z}) = \left(1 + |\mathbf{z}|^2\right) \widehat{\Phi}^2(\mathbf{z}).$$

From (Wendland 2005, Theorem 10.35), we know that $\widehat{\Phi}(\mathbf{z}) \geq C\|\mathbf{z}\|^{-2\tau}$ and hence $\widehat{\Upsilon}(\mathbf{z}) \geq C\|\mathbf{z}\|^{-4\tau}$. Then using (Wendland 2005, Theorem 12.3) we reach

$$\lambda_{\min}(\mathbf{G}) \geq C q_X^{4\tau - d}.$$

With the continuity of $a(\cdot,\cdot)$, (Wendland 1998*a*, Theorem 14.2) and the non-negativity of norms, we also have the following bound on the maximum eigenvalue

$$\begin{aligned} a(v,v) &\leq C\|v\|^2_{H^1(\Omega)} \leq C\|v\|^2_{H^1(\mathbb{R}^d)} = C\alpha^T \mathbf{G} \alpha \\ &\leq C N q_X^{-d} \|\alpha\|_2^2. \end{aligned}$$

These two bounds, in conjunction with (22), complete the proof. □

We will consider (23) further with the scaled RBFs $\Phi_i := \Phi_\delta(\cdot, \mathbf{x}_i)$, where Φ is the C^6 Wendland function given by

$$\Phi(\mathbf{x}) = (1 - \|\mathbf{x}\|)_+^8 \left(32\|\mathbf{x}\|^3 + 25\|\mathbf{x}\|^2 + 8\|\mathbf{x}\| + 1\right), \tag{24}$$

which is positive definite on $\mathbb{R}^2$ (Wendland 2005). Now since the support of the radial basis functions is fixed, $\|\Phi_i\|_{H^1(\mathbb{R}^d)}$ is fixed and is independent of the point set X and Ω and we can express this as

$$\|\Phi_i\|^2_{H^1(\mathbb{R}^d)} = \int_0^{2\pi} \int_0^{\delta} r \left(\phi_\delta^2(r) + \delta^2 \left(\frac{\mathrm{d}}{\mathrm{d}r} \phi_\delta(r) \right)^2 \right) \mathrm{d}r\, \mathrm{d}\theta. \tag{25}$$

In the case of the unit square, which will be used for the numerical experiments in Section 7, note that $\|\Phi_i\|_{H^1(\Omega)}$ is minimised when the RBF centre is located on a corner. This can be easily verified since moving the centre in any direction (in Ω) will keep the original area inside Ω and lead to additional area

being inside Ω and the integrand is non-negative. Then we can then bound $\|\Phi_i\|^2_{H^1(\Omega)}$ by

$$\|\Phi_i\|^2_{H^1(\Omega)} \geq \int_0^{\frac{\pi}{2}} \int_0^{\delta} r \left(\phi^2_\delta(r) + \delta^2 \left(\frac{\mathrm{d}}{\mathrm{d}r} \phi_\delta(r) \right)^2 \right) \mathrm{d}r \, \mathrm{d}\theta$$
$$= \frac{1}{4} \|\Phi_i\|^2_{H^1(\mathbb{R}^d)}, \tag{26}$$

where we have also used equations from (Fasshauer 2007, Appendix D).

This means that

$$\rho_i := \frac{\|\Phi_i\|_{H^1(\mathbb{R}^d)}}{\|\Phi_i\|_{H^1(\Omega)}} \leq 2.$$

As a result, if $\delta \leq q_X$, which means that there is no overlap between the various RBFs and $\mathbf{F}$ and $\mathbf{G}$ are diagonal, or if $\mathbf{F}$ and $\mathbf{G}$ are diagonally dominant, then (23) will hold since then we know that positive diagonal entries will ensure at least positive semi-definiteness. Whilst we have been unable to prove that (23) holds in full generality for the unit square, it is supported by extensive numerical testing. The numerical experiments in Section 7 also provide empirical evidence since Corollary 5.6 holds, which depends on Theorem 5.4.

We have the following theorem on the condition number of the multiscale algorithm.

Theorem 5.5 *Let Φ be a positive definite kernel generating $H^\tau(\mathbb{R}^d)$. Then the condition number of the Galerkin approximation matrices from Algorithm 1 can be bounded by*

$$\kappa(\mathbf{A}) \leq C \left(\frac{\delta}{q_X} \right)^{4\tau - 2d}, \tag{27}$$

with a constant $C > 0$ independent of X and of the scaling parameter δ.

Proof At each level, we now introduce the dataset $X/\delta = \{\mathbf{x_1}/\delta, \ldots, \mathbf{x}_M/\delta\}$, which obviously has separation radius

$$q_{X/\delta} = \frac{q_X}{\delta}$$

and since $a(\cdot,\cdot)$ is bilinear, the Galerkin approximation matrix at each level is

$$\mathbf{A}_{X,\delta} = (a(\Phi_\delta(\cdot, \mathbf{x_i}), \Phi_\delta(\cdot, \mathbf{x_j})))$$
$$= \left(\delta^{-2d} a \left(\Phi \left(\frac{\cdot - \mathbf{x}_i}{\delta} \right), \Phi \left(\frac{\cdot - \mathbf{x}_j}{\delta} \right) \right) \right) = \delta^{-2d} \mathbf{A}_{X/\delta,1}.$$

Then the result follows with Theorem 5.4. □

Corollary 5.6 *If the datasets are quasi-uniform, which means that h_j/q_j is bounded above by a constant, then the condition numbers of the Galerkin approximation matrices from Algorithm 1 are bounded above by a constant.*

Proof Algorithm 1 takes $\delta_j = \nu h_j$ with a constant $\nu > 1$. With the assumption of quasi-uniformity, $h_j \leq c q_j$ and the result follows with Theorem 5.5. □

We note that since we only require the bilinear form $a(\cdot,\cdot)$ to be continuous and coercive, these theorems on the condition number will apply for PDEs with Robin and/or Neumann boundary conditions as well as PDEs with Dirichlet boundary conditions.

6 A nested multiscale algorithm

In this section we will consider another multiscale Galerkin algorithm that was proposed in Wendland (1998*c*). This essentially extends Algorithm 1 and hence we can consider a PDE either with or without Dirichlet boundary conditions. We refer to this as a *nested* multiscale algorithm because it contains inner and outer iterations. We will also see that this has a connection to multigrid methods from the finite elements literature. The details are given in Algorithm 2.

Algorithm 2: Nested multiscale Galerkin approximation

Data: K: number of outer levels
n: number of inner levels
$\{X_\ell\}_{\ell=1}^n$: the set of centres for each inner level ℓ, with mesh norms at each inner level given by h_ℓ satisfying $c\mu h_\ell \leq h_{\ell+1} \leq \mu h_\ell$ with fixed $\mu \in (0,1), c \in (0,1]$ and h_1 sufficiently small
$\{\delta_\ell\}_{\ell=1}^n$: the scale parameters to use at each inner level, satisfying $\delta_\ell = \nu h_\ell$, ν a fixed constant.

begin
  Set $\widetilde{u}_0 = 0$
  for $k = 0, 1, 2, \dots, K$ **do**
    for $\ell = 1, 2, \dots, n$ **do**
      With the level-specific approximation subspace $V_\ell := \operatorname{span}\{\Phi_\ell(\cdot,\mathbf{x}), \mathbf{x} \in X_\ell\}$ solve for the Galerkin approximation given by

$$s_{kn+\ell} \in V_\ell \, : \, a(s_{kn+\ell}, v) = \langle f, v\rangle_{L_2(\Omega)} - a(\widetilde{u}_{kn+\ell-1}, v)$$

      for all $v \in V_\ell$. Update the solution according to

$$\widetilde{u}_{kn+\ell} = \widetilde{u}_{kn+\ell-1} + s_{kn+\ell}$$

Result: Approximate solution at level n, $\widetilde{u}_{n(K+1)}$
The error at level $n(K+1)$, $e_{n(K+1)} := u - \widetilde{u}_{n(K+1)}$.

From Wendland (1998*c*) we have the following theorem regarding convergence.

Theorem 6.1 *Let u^* denote the best approximation to u from $V_1 + \ldots + V_n$ with respect to the norm $\|\cdot\|_{H^1(\Omega)}$. Then there exists $\theta \in (0,1)$ such that*

$$\|u^* - \widetilde{u}_K\|_{H^1(\Omega)} \leq \theta^K \|u\|_{H^1(\Omega)},$$

where $\widetilde{u}_K$ is the approximation from Algorithm 2.

Note however that this does not mean that we have linear convergence of the approximation from Algorithm 2 to the true solution u. The convergence of the approximation from Algorithm 2 to the true solution u is given in the following theorem.

Theorem 6.2 *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary. Let Φ be a kernel generating $H^1(\mathbb{R}^d)$ and Φ_j be defined by (2) with scale factor δ_j. Then for Algorithm 2 there exist constants $\alpha_2 > 0, C_5 > 0, C_6 > 0$ such that*

$$\|u - \widetilde{u}_{n(K+1)}\|_{L_2(\Omega)} \leq C_4 \alpha_1^{n+K(n-1)} \alpha_2^K \|u\|_{H^1(\Omega)}, \quad K = 1, 2, \ldots$$

with α_1 given by (20).

Proof Since there are $K+1$ outer iterations (since the outer level index starts at 0) and n inner iterations, we have $(K+1)n$ iterations in total, of which $K(n-1)+n$ iterations are with subsequently decreasing scale parameters for which we can use Theorem 5.2. The remaining K iterations involve the subsequent error estimation for $K > 0$ and $\ell = 1$ since in this case, we have an increasing scale parameter for which Theorem 5.2 does not apply. With the proof of Theorem 5.2, we can derive a similar result with increasing scale parameters. In this case, we need to change the right hand side of (19) and the following line to

$$\left(\frac{\delta_{i+1}}{\delta_i}\right) \leq \mu^{-2},$$

and then we can define

$$\alpha_2 := \left(\nu^{-2} C_1/c_1 + \mu^{-2} C_2/c_1\right)^{1/2}. \tag{28}$$

Note that $\alpha_2 > \alpha_1$ since $\mu < 1$ by definition and the constants are all positive which will mean a lower rate of convergence as compared to Algorithm 1. The remainder of the proof follows the same steps as the proofs of Corollary 5.3 and we leave the details to the reader. □

Note that the original justification for proposing this nested multiscale algorithm was that the errors from Algorithm 1 appeared to be dominated by a global behaviour, suggesting the need to go back and fit on a coarse set of centres with a large support. This is a similar idea used in the multigrid method in the finite element literature (Brenner & Scott 2008, Chapter 6.3). As stated in (Fasshauer 2007, Chapter 44.3), this additional outer iteration is known from Kaczmarz iteration, which is frequently used in the multigrid literature as a smoother (McCormick 1992, Kang & Kwak 1997)).

7 Numerical experiments

In this section, we present the results from applying the multiscale and nested multiscale algorithms to various PDEs.

7.1 Multiscale Algorithm

In this subsection we consider two PDEs, the first without Dirichlet boundary conditions and the second with Dirichlet boundary conditions.

The first problem is the Helmholtz-like equation with natural boundary conditions:

$$\begin{aligned} -\Delta u + u &= f \quad \text{in} \quad \Omega, \\ \frac{\partial}{\partial \mathbf{n}} u &= 0 \quad \text{on} \quad \partial\Omega. \end{aligned}$$

We take $\Omega = [-1,1]^2$ and $f(x,y) = \cos(\pi x)\cos(\pi y)$. The outer unit normal vector is denoted by $\mathbf{n}$. The exact solution is given by

$$u(x,y) = \frac{\cos(\pi x)\cos(\pi y)}{2\pi^2 + 1}.$$

We again use the C^6 Wendland radial basis function given by (24). We used five levels for the approximation, with equally spaced point sets at each level. The number of points, N, and the fill distances, h, are given in Table 1. We note that the fill distances decrease by almost exactly one half at each level and hence we select $\mu = \frac{1}{2}$. The L_2 and L_∞ errors and condition numbers (κ)

Level	1	2	3	4	5
N	25	81	289	1089	4225
h	3.5e-1	1.75e-1	8.75e-2	4.37e-2	2.19e-2

Table 1: The number of uniform points used at each level and the associated fill distance for the multiscale Galerkin approximation example

of the stiffness matrix are given in Table 2. The L_2 error was estimated using Gaussian quadrature with a 300×300 tensor product grid of Gauss-Lobatto points and the L_∞ error was estimated with the same tensor product grid.

The second example uses the Poisson problem

$$\begin{aligned} -\Delta u &= f \quad \text{in} \quad \Omega, \\ u &= 0 \quad \text{on} \quad \partial\Omega. \end{aligned}$$

We take $\Omega = [-1,1]^2$ and $f(x,y) = \sin(\pi x)\cos\left(\frac{\pi}{2}y\right)$. The exact solution is given by

$$u(x,y) = \frac{\sin(\pi x)\cos\left(\frac{\pi}{2}y\right)}{1.25\pi^2}.$$

Level	1	2	3	4	5
δ_j	2	1	0.5	0.25	0.125
$\|e_j\|_2$	8.000e-4	2.145e-4	1.059e-4	7.009e-5	5.178e-5
$\|e_j\|_\infty$	1.721e-3	7.273e-4	3.764e-4	2.147e-4	1.394e-4
κ_j	1.608e+3	3.125e+3	4.159e+3	4.576e+3	4.710e+3

Table 2: The scaling factors, approximation errors and condition numbers of the stiffness matrices for the multiscale Galerkin algorithm with a Neumann boundary condition.

We again use the C^6 Wendland function as the kernel, with the same 5 levels as for the first example. To verify that (15) holds, we first check for the basis functions. For the boundary norm, since $\Omega = [-1,1]^2$, without loss of generality, we consider the $x = -1$ boundary of the domain only. Then we have boundary integrals of the form (Fasshauer 2007, Appendix D)

$$\begin{aligned}
\|\nabla\Phi_\delta \cdot \mathbf{n}\|^2_{L_2(\partial\Omega)} &= \int_{-\delta}^{\delta} \left(\frac{\partial \phi_\delta}{\partial y}\right)^2 \mathrm{d}y \\
&= \delta^2 \int_{-\delta}^{\delta} \left(\frac{22y}{\delta^2}\left(\frac{16y^2}{\delta^2} + \frac{7y}{\delta} + 1\right)\left(1 - \frac{y}{\delta}\right)^7\right)^2 \mathrm{d}y \\
&= \frac{603969552384\delta}{11305}.
\end{aligned}$$

For the interior, we have

$$\begin{aligned}
\|\nabla\Phi_\delta\|^2_{L_2(\Omega)} &= \delta^2 \int_{\Theta}\int_0^{\delta} r\left(\frac{d}{\mathrm{d}r}\phi_\delta(r)\right)^2 \mathrm{d}r\,\mathrm{d}\theta \\
&= \delta^2 \int_{\Theta}\int_0^{\delta} r\left(\frac{22r}{\delta^2}\left(\frac{16r^2}{\delta^2} + \frac{7r}{\delta} + 1\right)\left(1 - \frac{r}{\delta}\right)^7\right)^2 \mathrm{d}r\,\mathrm{d}\theta \\
&= \frac{2453\delta^2}{4845}\int_{\Theta} \mathrm{d}\theta,
\end{aligned}$$

where Θ specifies the support of ϕ in Ω and hence this last expression is finite and does not depend on δ.

In practical applications, we need to select a value of β_N satisfying $\beta_N > 2C_N^2/\delta$. In Griebel & Schweitzer (2002), it is proposed to estimate $C_N/\sqrt{\delta}$ as the maximum eigenvalue of the generalised eigenvalue problem,

$$\mathbf{B}\mathbf{v} = \lambda\mathbf{D}\mathbf{v}, \tag{29}$$

where

$$B_{ij} = \int_{\partial\Omega} (\nabla\Phi_i \cdot \mathbf{n})(\nabla\Phi_j \cdot \mathbf{n}), \tag{30}$$

and

$$D_{ij} = \int_{\Omega} \nabla \Phi_i \cdot \nabla \Phi_j, \tag{31}$$

where i and j run over the indices of all the radial basis functions with support overlapping the boundary. The extra calculation involved in this step is not significant since the entries of $\mathbf{B}$ are required for the construction of the stiffness matrix and the set of centres overlapping the boundary will generally be small compared to the entire set of centres. The maximum eigenvalue can also be efficiently computed with a simultaneous Rayleigh-quotient minimisation method (Longsine & McCormick 1980).

Level	1	2	3	4	5
δ_j	2	1	0.5	0.25	0.125
$\|e_j\|_2$	8.125e-3	1.451e-3	3.229e-4	8.217e-5	2.219e-5
$\|e_j\|_\infty$	1.057e-2	2.350e-3	6.504e-4	1.963e-4	5.962e-5
κ_j	5.633e+5	1.001e+6	8.056e+5	4.572e+5	2.374e+5

Table 3: The scaling factors, approximation errors and condition numbers of the stiffness matrices for the multiscale Galerkin algorithm with Dirichlet boundary conditions

The results are in Table 3 and support the theoretical findings above. We note that whilst Wendland (1998*c*) did not find convergence after the third level with a similar algorithm, this may be due to the approximations used to calculate the integrals, rather than the algorithm itself. The potential for errors in integration to affect the performance of Galerkin techniques are well known (Strang 1973, Brenner & Scott 2008). To estimate the integrals, we used the MATLAB functions `quad2d` and `quad` with an absolute tolerance value of $1e^{-10}$. We also estimated the non-zero integration range both to speed up the calculations as well as to reduce numerical error which can result if for example, we integrate over the entire domain $[-1, 1]^2$ whilst the function only has a very small support.

7.2 Nested multiscale algorithm

In this subsection, we consider the same example as in Section 7.1 however now with Algorithm 2 with $K = 2$ and $n = 2$. We use the first two levels of the example described in Section 7.1 as the inner iteration. We also use the same kernel. Our choice of K and n leads to a 6 level algorithm. A similar example was considered in (Wendland 1998*c*, Section 5), however a lack of information regarding the exact approximation spaces used for the inner and outer level iterations means we have not been able to compare our results. The results from this 6 level nested algorithm are in Table 4.

Level	1	2	3	4	5	6
N	25	81	25	81	25	81
$\|e_j\|_2$	8.000e-4	2.145e-4	2.045e-4	2.088e-4	1.991e-4	2.034e-4
$\|e_j\|_\infty$	1.721e-3	7.274e-4	4.325e-4	7.002e-4	4.165e-4	6.742e-4
κ_j	1.608e+3	3.125e+3	1.608e+3	3.125e+3	1.608e+3	3.125e+3

Table 4: The number of centres, scaling factors, approximation errors and condition numbers of the stiffness matrices for the nested multiscale Galerkin algorithm

The results indicate erratic convergence and approximation errors far inferior to those using Algorithm 1. This is not surprising since Theorem 8.2 indicates convergence of our approximation to the best approximation to u from $V_1 + V_2$ whilst in Algorithm 1 our approximation is formed from $V_1 + \ldots + V_5$.

8 Analysis of convergence

In this section we will focus on estimation of the convergence and verifying approximation orders. Similarly to Wendland (2010), we will also rewrite the convergence results in terms of fill distances, which is the usual form of convergence results for radial basis function approximations (Wendland 2005, Fasshauer 2007).

8.1 Multiscale Galerkin algorithm

We consider Algorithm 1 with $h_1 = \mu$ and $h_{j+1} = \mu h_j$. Since μ is a constant, we can rewrite (20) as

$$\alpha_1 = c_3 \mu.$$

Then with Corollary 5.3 we have that

$$\|e_n\|_{L_2(\Omega)} := \|u - \widetilde{u}_n\|_{L_2(\Omega)} \leq C h_n^{1-\sigma} \|u\|_{H^1(\Omega)}, \tag{32}$$

with

$$\sigma := -\log c_3 / \log \mu. \tag{33}$$

Hence we can either express our convergence in terms of an exponent of h_n or equivalently α_1^n.

It is of interest that the error bounds do not depend on the kernel used for the approximation spaces. Typically with a kernel which generates $H^\tau(\mathbb{R}^d)$, we see error bounds proportional to h^τ. Since our kernel for the error analysis generates $H^1(\mathbb{R}^d)$, we have h_n^1. Henceforth, we analyse the convergence in terms of α_1. We can calculate estimates of α_1, which we denote by $\widetilde{\alpha}_1$, as follows

$$\widetilde{\alpha}_{1,n} := \frac{\|e_n\|_{L_2(\Omega)}}{\|e_{n-1}\|_{L_2(\Omega)}}.$$

	C^2 WF	C^6 WF
$\widetilde{\alpha}_{1,2}$	0.128	0.268
$\widetilde{\alpha}_{1,3}$	0.318	0.494
$\widetilde{\alpha}_{1,4}$	0.507	0.662
$\widetilde{\alpha}_{1,5}$	0.618	0.739

Table 5: The estimated convergence rates $\widetilde{\alpha}_{1,n}$ using the results for the L_2 norm errors from the first example in Section 7.1 with the C^2 and C^6 Wendland functions.

We can see that the estimated values of α_1 are higher with the C^6 Wendland function, which indicates that we should not necessarily expect faster convergence with a smoother Wendland function and consequently we should not expect an error bound proportional to h^τ.

8.2 Nested multiscale Galerkin algorithm

In this subsection, we will focus on considering convergence of the nested multiscale Galerkin algorithm in terms of α_1 and α_2. Note that a bound for the error at level $n(K+1)$ in terms of the fill distance at level $n(K+1)$, such as in (32), will not be possible here because the fill distance at level $n(K+1)$ only depends on n and not on K. In other words, increasing K has no effect on the final fill distance.

An additional benefit of considering the nested multiscale algorithm is more estimates of α_1, particularly for repeated applications of the inner iterations (when $K > 1$). Table 6 gives the estimates of α_1 and α_2 from considering successive L_2 norm error estimates in Section 7.2. Successive error estimates will be of the form

$$\frac{\|e_i\|_{L_2(\Omega)}}{\|e_{i-1}\|_{L_2(\Omega)}}.$$

By definition of our nested multiscale algorithm, for $i = n+1, 2n+1, \ldots$ we have an estimate for α_2 and in all other cases, an estimate for α_1.

Table 7.2 presents the estimated convergence rates $\widetilde{\alpha}_{1,n}$ and $\widetilde{\alpha}_{2,n}$ using the results for the L_2 norm errors from the example in Section 7.2.

Level	$\widetilde{\alpha}_{1,n}$	$\widetilde{\alpha}_{2,n}$
2	0.268	
3		0.953
4	1.021	
5		0.954
6	1.022	

Table 6: The estimated convergence rates $\widetilde{\alpha}_{1,n}$ and $\widetilde{\alpha}_{2,n}$ using the results for the L_2 norm errors from the example in Section 7.2.

Interestingly, the difficulties with convergence are not due to α_2 which is seen to be less than 1 in all cases. This is empirical evidence of the effectiveness of the smoothing nature of the inner iterations, in that after the inner iterations, the errors are again of a global nature and hence a return to a coarse grid is justified. Convergence is affected by the repeated application of the inner iterations for which $\widetilde{\alpha}_{1,n}$ is always greater than 1. Empirically, this appears to suggest that the more localised features have already been captured in the approximate solution.

This also implies that the angles between the approximation subspaces are close to zero since the linear convergence rate in Theorem 6.1 can be bounded above by the angle between the subspaces. This is covered next.

8.2.1 Angles between subspaces

For our analysis of the convergence of the nested multiscale algorithm, we require the following definition of the angle between subspaces.

Definition 8.1 *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be closed subspaces of a Hilbert space $\mathcal{H}$ with $U := \mathcal{H}_1 \cap \mathcal{H}_2$. Then the angle α between $\mathcal{H}_1$ and $\mathcal{H}_2$ is given by*

$$\cos\alpha = \sup\left\{\langle u, v\rangle : u \in \mathcal{H}_1 \cap U^{\perp}, v \in \mathcal{H}_2 \cap U^{\perp} \text{ and } \|u\|, \|v\| \le 1\right\}.$$

It is well known (Smith et al. 1977, Deutsch 1990) that Algorithm 2 converges linearly in the following sense.

Theorem 8.2 *Let u^* be the best approximation to u from $V_1 + \ldots + V_j$ with respect to $\|\cdot\|_{H^1(\Omega)}$. Let $\widetilde{u}$ be the approximation from Algorithm 2. Let α_j be the angle between V_j and $\cap_{i=j+1}^{k} V_i$. Then*

$$\|u^* - \widetilde{u}\|_{H^1(\Omega)} \le c^K \|u\|_{H^1(\Omega)},$$

where

$$c^2 \le 1 - \prod_{j=1}^{n-1} \sin^2\alpha_j.$$

This means that we need to estimate $\sin\alpha_j$ to obtain upper bounds for the convergence rate. We follow a similar approach to Beatson et al. (2000). We firstly define modified sets of centres as $\widetilde{X}_1 = X_1$ and

$$\widetilde{X}_i = X_i \setminus \bigcup_{j=1}^{i-1} X_j, \quad i \ge 2,$$

and the corresponding approximation spaces as

$$\widetilde{V}_i = \operatorname{span}\left\{\Phi_{\delta_i}(\cdot, \mathbf{x}), \mathbf{x} \in \widetilde{X}_i\right\}.$$

Then we need to find the supremum of the inner product of $u \in \widetilde{V}_i$ and $v \in \mathcal{A}_{i+1}$ where

$$\mathcal{A}_{i+1} = \bigcup_{j=i+1}^{K} \widetilde{V}_j = \operatorname{span}\left\{ \Phi_{\delta_j}(\cdot, \mathbf{x}_j), \mathbf{x}_j \in \bigcup_{j=i+1}^{K} \widetilde{X}_j \right\},$$

with $\|u\| = \|v\| = 1$. With the matrix $\mathbf{K}^{\{12\}}$ given by

$$K_{i,j}^{\{12\}} = \left\langle \Phi_{\delta_i}(\cdot, \mathbf{x}_i), \Phi_{\delta_j}(\cdot, \mathbf{x}_j) \right\rangle, \mathbf{x}_i \in \widetilde{X}_i, \mathbf{x}_j \in \bigcup_{j=i+1}^{K} \widetilde{X}_j,$$

and with coefficient vectors μ and ν for u and v respectively, we seek the supremum of $\mu \mathbf{K}^{\{12\}} \nu$. We also define matrices $\mathbf{K}^{\{1\}}$ and $\mathbf{K}^{\{2\}}$ as

$$K_{ij}^{\{1\}} = \left\langle \Phi_{\delta_i}(\cdot, \mathbf{x}_i), \Phi_{\delta_j}(\cdot, \mathbf{x}_j) \right\rangle, \mathbf{x}_i \in \widetilde{X}_i, \mathbf{x}_j \in \widetilde{X}_i,$$

and

$$K_{ij}^{\{2\}} = \left\langle \Phi_{\delta_i}(\cdot, \mathbf{x}_i), \Phi_{\delta_j}(\cdot, \mathbf{x}_j) \right\rangle, \mathbf{x}_i \in \bigcup_{j=i+1}^{K} \widetilde{X}_j, \mathbf{x}_j \in \bigcup_{j=i+1}^{K} \widetilde{X}_j.$$

Let the Cholesky decomposition of $\mathbf{K}^{\{1\}}$ be $\mathbf{L}_1^T \mathbf{L}_1$. This is well-defined since $\mathbf{K}^{\{1\}}$ is strictly positive definite and symmetric. Then $\|u\|^2 = \mu^T \mathbf{L}_1^T \mathbf{L}_1 \mu$ and letting $\boldsymbol{\gamma}_1 = \mathbf{L}_1 \mu$ gives $\|u\|^2 = \boldsymbol{\gamma}_1^T \boldsymbol{\gamma}_1$. We can follow a similar approach with $\mathbf{K}^{\{2\}}$ which gives $\|v\|^2 = \boldsymbol{\gamma}_2^T \boldsymbol{\gamma}_2$ with $\mathbf{K}^{\{2\}} = \mathbf{L}_2^T \mathbf{L}_2$ and $\boldsymbol{\gamma}_2 = \mathbf{L}_2 \nu$. However in this case since $\mathbf{K}^{\{2\}}$ is the union of radial basis functions with (possibly) different scaling factors, we cannot be sure that $\mathbf{K}^{\{2\}}$ is positive definite. In our example, $\mathbf{K}^{\{2\}}$ was always positive definite and we do not dwell further on this. Sufficient conditions for an interpolation matrix constructed with several scaling factors to be positive definite can be found in (Bozzini et al. 2004, (11)) which also requires the Fourier transform of a Wendland function from Chernih et al. (2011) to compute a lower bound on the minimum eigenvalue as given in (Wendland 2005, Theorem 12.3).

Then we have

$$\begin{aligned} \langle u, v \rangle &= \mu^T \mathbf{K}^{\{12\}} \nu \\ &= \mu^T \mathbf{L}_1^T (\mathbf{L}_1^{-1})^T \mathbf{K}^{\{12\}} \mathbf{L}_2^{-1} \mathbf{L}_2 \nu \\ &= \boldsymbol{\gamma}_1^T \mathbf{M} \boldsymbol{\gamma}_2, \end{aligned}$$

with $\mathbf{M} := (\mathbf{L}_1^{-1})^T \mathbf{K}^{\{12\}} \mathbf{L}_2^{-1}$. The supremum of the inner product is given by the largest singular value of $\mathbf{M}$. We denote this supremum by $\sin \widetilde{\alpha}_j$ and the results with $\{\widetilde{V}_i\}_{i=1}^5$ are in Table 7. We note that since $\widetilde{X}_i \subseteq X_i$, $\sin \widetilde{\alpha}_i$ is a lower bound on $\sin \alpha_i$. We chose to estimate $\sin \widetilde{\alpha}_i$ since by removing nested centres from later levels, we had less difficulties with singular matrices.

i	1	2	3	4
$\sin\widetilde{\alpha}_i$	9.849e-3	2.684e-2	4.153e-2	6.987e-2

Table 7: The estimates of $\sin\widetilde{\alpha}_i$ with the approximation spaces $\{\widetilde{V}_i\}_{i=1}^5$.

9 Acknowledgments

We used some datasets and code from Fasshauer (2007) as well as Andrea Tagliasacchi's "kd-tree for matlab" library, available at `www.mathworks.com/matlabcentral/fileexchange/21512`. This work was supported by the Australian Research Council. The authors thank Ian H. Sloan and Robert S. Womersley for helpful discussions.

References

Beatson, R. K., Light, W. A. & Billings, S. (2000), 'Fast solution of the radial basis function interpolation equations: Domain decomposition methods', *SIAM J. Sci. Comput.* **22**, 1717–1740.

Blowey, J., Craig, A. & Shardlow, T. (2003), *Frontiers in Numerical Analysis*, Universitext, Springer.

Bozzini, M., Lenarduzzi, L., Rossini, M. & Schaback, R. (2004), 'Interpolation by basis functions of different scales and shapes', *Calcolo* **41**, 77–87.

Brenner, S. C. & Scott, L. R. (2008), *The Mathematical Theory of Finite Element Methods*, Texts in Applied Mathematics, 3 edn, Springer.

Chernih, A., Sloan, I. H. & Womersley, R. S. (2011), Wendland functions with increasing smoothness converge to a Gaussian.

Deutsch, F. (1990), The method of alternating orthogonal projections, *in* S. P. Singh, ed., 'Approximation Theory, Spline Functions and Applications', Kluwer Academic Publishers, pp. 105–121.

Fasshauer, G. (1999), 'Solving differential equations with radial basis functions: multilevel methods and smoothing', *Advances in Computational Mathematics* **11**, 139–159.

Fasshauer, G. E. (2007), *Meshfree Approximation Methods with MATLAB*, Vol. 6 of *Interdisciplinary Mathematical Sciences*, World Scientific Publishing Co., Singapore.

Floater, M. S. & Iske, A. (1996), 'Multistep scattered data interpolation using compactly supported radial basis functions', *Journal of Computational and Applied Mathematics* **73**, 65–78.

Giesl, P. & Wendland, H. (2007), 'Meshless collocation: Error estimates with application to dynamical systems', *SIAM Journal on Numerical Analysis* **45**, 1723–1741.

Griebel, M. & Schweitzer, M. (2002), A particle-partition of unity method. part v: boundary conditions, *in* S. Hildebrandt & H. Karcher, eds, 'Geometric Analysis and Nonlinear Partial Differential Equations', Springer, Berlin, pp. 519–542.

Kang, K. S. & Kwak, D. Y. (1997), Convergence estimates for multigrid algorithms with Kaczmarz smoothing, *in* R. Glowinski, J. Périaux, Z.-C. Shi & O. Widlund, eds, '8th International Conference on Domain Decomposition Methods', John Wiley and Sons.

Le Gia, Q. T., Sloan, I. H. & Wendland, H. (2010), 'Multiscale analysis in Sobolev spaces on the sphere', *SIAM Journal on Numerical Analysis* **48**, 2065–2090.

Longsine, D. E. & McCormick, S. F. (1980), 'Simultaneous Rayleigh-quotient minimization methods for $ax = \lambda bx$', *Lin. Alg. Appl* **34**, 195–234.

Maz'ya, V. (2011), *Sobolev Spaces: with Applications to Elliptic Partial Differential Equations*, Vol. 342 of *Grundlehren der mathematischen Wissenschaften*, 2 edn, Springer.

McCormick, S. F. (1992), *Multilevel Projection Methods for Partial Differential Equations*, Vol. 62 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia.

Nitsche, J. (1970–1971), 'Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind', *Abh. Math. Univ. Hamburg* **36**, 9–15.

Schaback, R. (1996), On the efficiency of interpolation by radial basis functions, *in* C. R. A. Le Méhauté & L. L. Schumaker, eds, 'Surface Fitting and Multiresolution Methods', Vanderbilt University Press, pp. 309–318.

Smith, B., Bjorstad, P. & Gropp, W. (1996), *Domain Decomposition, Parallel Multilevel Methods for Elliptic Partial Differential Equations*, Cambridge University Press.

Smith, K. T., Solmon, D. C. & Wagner, S. L. (1977), 'Practical and mathematical aspects of the problem of reconstructing objects from radiographs', *Bulletin of the American Mathematical Society* **83**, 1227–1270.

Stein, E. M. (1970), *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, Princeton, New Jersey.

Strang, G. (1973), 'Piecewise polynomials and the finite element method', *Bull. Amer. Math. Soc.* **79**, 1128–1137.

Weinan, E., Engquist, B., Li, X., Ren, W. & Vanden-Eijnden, E. (2007), 'Heterogeneous multiscale methods: a review', *Communications in Computational Physics* **2**, 367–450.

Wendland, H. (1998*a*), 'Error estimates for interpolation by compactly supported radial basis functions of minimal degree', *Journal of Approximation Theory* **93**, 258–272.

Wendland, H. (1998*b*), 'Meshless Galerkin methods using radial basis functions', *Mathematics of Computation* **68**(228), 1521–131.

Wendland, H. (1998*c*), Numerical solution of variational problems by radial basis functions, *in* C. K. Chui & L. L. Schumaker, eds, 'Approximation Theory IX, Volume 2: Computational Aspects', Vanderbilt University Press, pp. 361–368.

Wendland, H. (2005), *Scattered Data Approximation*, Vol. 17 of *Cambridge Monographs on Applied and Computational Mathematics*, Cambridge University Press, Cambridge.

Wendland, H. (2010), 'Multiscale analysis in Sobolev spaces on bounded domains', *Numerische Mathematik* **116**, 493–517.